DOCUMENT RESUME

ED 206 362 CS 006 309

AUTHOR Sakiey, Elizabeth
TITLE Word and Syllable Counts in Theoretical and Practical Perspectives.
PUB DATE Dec 80
NOTE 12p.; Paper presented at the Annual Meeting of the National Reading Conference (30th, San Diego, CA, December 3-6, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computational Linguistics; *Reading Research; *Research Needs; *Structural Analysis (Linguistics); *Syllables; Word Frequency; Word Lists
IDENTIFIERS *Syllabaries

ABSTRACT

In contrast to research and application of word frequency counts, there has been little or no use of syllable frequency counts. The few syllable counts that have been conducted are not well known, not often used, and not easily obtained. This is unfortunate, because a publicized, easily obtainable, computer syllable analysis would have numerous applications, including the following: (1) serving as a basis for a decoding strategy; (2) influencing the selection of words used in basal and remedial reading materials, trade books, and spelling, typing, stenography, and other instructional materials related to business education; (3) providing a database for designing readability formulas, auditory perception tests, tachistoscopic report studies, and verbal learning and linguistic experiments; and (4) functioning as a component in a reading model. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208362

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Thirtieth Annual Meeting

of the National Reading Conference

San Diego, CA

December 3-6, 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Elizabeth Sakiey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

WORD AND SYLLABLE COUNTS IN THEORETICAL AND PRACTICAL PERSPECTIVES

Elizabeth Sakiey

Glassboro State College

Frequency counts are building blocks in the construction of strategies and materials for teaching reading. For that reason, they should be taken into account when constructing a model of reading. Unfortunately, there are few available syllable counts that are appropriate for developing theories, research designs, materials, and instructional strategies.

Frequency Counts in Theoretical Perspective

Several reading models have incorporated the common observation that readers process frequent words more easily than rarely encountered ones. Venezsky and Calfee (1970) contend that very common words are retrived from a highly organized store that can be searched quickly on the

CS006309

basis of such features as initial letters and length. Less common words, on the other hand, are more accessible by sound because they are stored differently. It seems logical to assume that frequently encountered syllables are also more efficiently processed. The LaBerge and Samuels (1974) model also focuses on the degree of automaticity of word processing. They propose that very common words are automatically coded into a visual word code which excites a meaning code. Unfamiliar words are automatically coded into spelling patterns (which are sometimes syllables) and phonological codes.

Moreover, studies indicate that there is a theoretical base for also including syllables in a model of reading. Instructional materials for beginning readers constantly refer to "words," "letters," "sounds," and "sentences" taking for granted that children are aware of these concepts. Often they are not. Smith and Miller (1966) have shown that Russian children from three to seven years of age are unable to divide simple sentences into word units or lexical constituents. These findings have been replicated and extended with English speaking children (Chappell, 1968; MacGinitie & Holden, 1972). In addition, there is evidence that syllable segmentation is easier than phoneme segmentation for primary grade children (Elkonin, 1973; Gibson & Levin, 1975; Gleitman & Rozin, 1973; Leong & Haines, 1978). Also Fox and Routh (1975) have noted a developmental progression in the ability to analyze spoken sentences into words, words into syllables, and syllables into speech sounds in children aged three to seven. Gibson and Levin (1975) also discuss a multilevel approach to reading involving the use of sentences, words, syllables, and letter clusters. Although they provide evidence for instruction in intraword relations for generalization to new words,

they do not specify the best method for such analysis.

Frequency Counts in Practical Perspective

Reading instruction has been greatly influenced by the major word frequency counts. Basal reader vocabularies, sight word lists for developmental as well as remedial instruction, and readability formulas, for the most part, have been dependent upon a few major counts. However, other lesser known counts have also made a contribution.

Word Counts

The most influential and well-known count of written materials was the Teacher's Word Book of 30,000 Words by Thorndike and Lorge (1944). It represents the most extensive sampling of print to date, approximately 18 million words from materials of both children and adults. It is gradually being replaced by two computerized lists: Computational Analysis of Present Day American English (Kucera-Francis, 1967) and the American Heritage Word Frequency Book (Carroll, Davies, & Richman, 1971). The corpus of the Kucera-Francis List is a one million running-word sample taken from 15 different kinds of adult reading materials. The 50,406 types (different words) on the list are presented in both alphabetical and rank order. In contrast, the 86,741 word American Heritage List, which is similar in construction, is based on a corpus of over 5 million running words taken from the printed vocabulary to which students in grades 3 to 9 are exposed. Frequencies for each word are provided for each grade level and for each of the 17 subject areas sampled.

These large computer based counts are inconvenient, difficult, and too expensive for the classroom teacher to use. Thus a number of shorter word lists, ranging in size from 300 to 3,000 have been

developed from them (Fry, 1980; Jonnson, 1971; Otto & Chester, 1972; Sakiey & Fry, 1979; Walker, 1979).

During the seventies, several other computational analyses were done on materials found in elementary classrooms. The Harris-Jacobson List (1972) contains words from 14 series of textbooks for grades 1 to 6 which yielded a corpus of 4 and a half million running words. Unique in its focus on two- and three-word strings rather than individual words, Hopkins and Moe (1978) examined the complete text, 202,763 running words, of 250 children's trade books in primary grades.

Counts have also been made of writing samples (Hillerich, 1978; Horn, 1926; Rinsland, 1945). But they are not limited solely to writtem materials. DeRocher, Miron, Patten, and Pratt (1973) list numerous oral word counts.

Syllable Counts

Contrasted to word frequency counts, the history of syllable counts is sparse. Prior to 1970, only three syllable counts had been completed. Dewey, in 1923, analyzed 100,000 running words on the adult level and "foneticaily" transcribed them. Of the 143,000 syllable occurrences, the 220 most common were listed in order of frequency. His 1950 expansion of the work included data on initial, medial, and final occurrences of these phonetic syllables. In 1938, Dolch found 8,509 different syllables in a sampling of 14,000 running words in elementary texts. From this data came a list of the 100 most frequently occurring syllables. It was not until 1954 than an inventory of initial, medial, and final syllables was completed by Osburn. Using the Rinsland List (1945), he discovered that there were 23,000 syllables in approximately 9,000 polysyllabic words.

None of these lists is appropriate for most instructional or research needs. Dewey's phonetic "sillables" were gathered from adult materials of the early nineteen hundreds and is therefore outdated. Since the list is not in graphemic form, it is sometimes difficult to utilize. The Dolch List, while graphemic, lists the 100 most common syllables in alphabetical form only, not in rank order and is based on elementary texts of the 1930's. The list, furthermore, is unknown, unused, and not easily obtained. The Osburn inventory is also somewhat out of date since it was based on the Rinsland List whose corpus was derived from children's writings of the 1930's. Moreover, the Osburn study was limited to polysyllabic words even though nearly half of all words are monosyllabic (Zipf, 1935). It, too, is difficult to locate.

The first computer analysis of syllable frequency was done by Sakiey (1979) who examined the 5,000 most frequent words of an edited version of the American Heritage List. There were 9,358 syllable occurrences with frequency ranges of 1 to 230. Of these 3,402 were different syllables. The 322 Syllables with frequencies over five were listed in alphabetical and descending frequency order. Since the frequency of each syllable was equal to the number of different words in which it appeared, it was termed an unweighted count. Then using the same data base, the entire corpus of unweighted syllables was merged with the frequencies of the words in which they occurred. This resulted in a 290-item weighted syllable list (Sakiey, Fry, Goss, & Loigman, in press). Weighting of the syllables was accomplished by summing the frequencies of the various words in which the syllables appeared. In other words, a weighted syllable equals the sum of the frequencies of all the words in which it occurs.

In extending this work, Sakiey and Martin (1980) investigated the extent to which Sakiey's unweighted, common syllables (1979) were present in the vocabulary introduced in six leading basal series in the primary grades. The words contained 80 percent of the common syllables, and in all the words taught, 60 percent of the syllables had two or less pronunciations. Most, 66 percent, had only one pronunciation.

Syllables as a decoding strategy. Syllable recognition as a teaching strategy has met with some success. Jones (1970) used closed syllables, termed "graphonemes," to improve the reading skills of Alaskan children. Gleitman and Rozin in 1973 demonstrated that it is possible to teach kindergarteners a syllabary. Their subjects were successful in combining syllables into words when shown pictures that represented syllables. Cunningham (1979) trained her fourth and fifth grade subjects to decode unfamiliar two-syllable words by comparing and contrasting the parts to known words and word parts, which were frequently common syllables. A similar method was also utilized by Martin and Sakiey (1979) with two mentally retarded students. Their sight vocabularies were increased considerably when taught the pronunciation and blending of selected common syllables (Sakiey, 1979).

Discussion

Word counts have been the subject of renewed interest in recent years, while syllables still receive little attention. This is, in part, because word counts function so many different ways, such as: (1) serving as a basis for dictionaries, basal reader vocabularies, and other instructional materials in reading, spelling, and business subjects, (2) generating new, short word lists for teaching sight words, (3) being a factor in new methods

of estimating readability, (4) providing a data base for developing vocabulary tests and verbal learning experiments, and (5) functioning as a component in reading models (LaBerge & Samuels, 1974; Venezsky & Calfee, 1970). But with the exception of the work of Martin and Sakiey (1979; Sakiey & Martin, 1980) there has been little or no utilization of the syllable counts described. As noted above, most of the syllable counts are not readily attainable, nor are they well known. A publicized, easily obtainable, modern syllable count would have numerous applications, including: (1) serving as a basis for a decoding strategy, such as the one developed by Martin and Sakiey (1979), (2) influencing the selection of words used in basal and remedial reading materials, trade books, as well as spelling, typing, and stenography materials, (3) providing a data base for designing readability formulas, auditory perception tests, tachistoscopic report studies, and verbal learning and linguistic experiments, and (4) functioning as a component in a reading model.

Conclusion

More research is needed in the area of frequency counts, particularly for the syllable. As more investigations are conducted, less ambiguity will surround the subject of syllables. As that happens, theorists will be more likely to include the syllable in their models of reading, practitioners will be more inclined to teach syllables by sight along with their most common pronunciations, and educational writers will incorporate high frequency syllables into beginning and remedial reading materials.

Eventually continued research may lead to the syllable's acceptance as a viable decoding strategy. At present, it is overshadowed by the attention given other units, the word and the letter. Though perhaps one day, the

phrase "Basic Sight Syllabes" may become as commonplace as "Basic Sight Vocabulary." But this will occur only after the data base on syllable frequency is extended.

References

Carroll, J. B., Davies, P. & Richman, B. American heritage word frequency book. Boston: Houghton Mifflin, 1971.

Chappell, P. R. Early development of awareness of lexical units. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1968.

Cunningham, P. M. A compare/contrast theory of mediated word identification. Reading Teacher, 1979, 32, 774-778.

DeRocher, J. E., Miron, M. S., Patten, S. M., & Pratt, C. C. The counting of words. Syracuse, NY: Syracuse University Research Corporation, 1973.

Dewey, G. Relativ frequency of English speech sounds. Cambridge: Harvard University Press, 1950. (Originally published, 1923.)

Dolch, E. W. Phonics and polysyllables. Elementary English Review, 1938, 15, 120-124.

Elkonin, D. B. U.S.S.R. In J. Downing (Ed.), Comparative reading: Cross-national studies of behavior and processes in reading and writing. New York: Macmillan, 1973.

Fox, B. & Routh, D. B. Phonetic analysis and synthesis as word-attack skills. Journal of Educational Psychology, 1976, 68, 70-74.

Fry, E. The new instant word list. Reading Teacher, 1980, 34, 284-289.

Gibson, E. J. & Levin, H. The psychology of reading. Cambridge: MIT Press, 1975.

Gleitman, L. R. & Rozin, P. Teaching reading by use of a syllabary. Reading Research Quarterly, 1973, 8, 447-483.

Harris, A. & Jacobson, M. Basic elementary reading vocabularies. New York: Macmillan, 1972.

Millerich, R. L. A writing vocabulary of elementary children. Springfield, IL: Charles C. Thomas, 1978.

Hopkins, C. J. & Moe, A. J. The computer assisted identification of common word strings from the text of children's trade books in primary grades. Paper presented at the meeting of the International Reading Association, Houston, May 1978.

Horn, E. A basic writing vocabulary. Iowa City, IA: College of Education, University of Iowa, 1926.

Johnson, D. The Dolch list re-examined. Reading Teacher, 1971, 24, 449-457.

Jones, V. W. Decoding and learning to read. Portland, OR: Northwest Regional Educational Laboratory, 1970.

Kucera, H. & Francis, W. N. Computational analysis of present-day American English. Providence, RI: Brown University Press, 1967.

LaBerge, D. & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Leong, C. K. & Haines, C. F. Beginning readers' analysis of words and sentences. Journal of Reading Behavior, 1978, 10, 393-407.

MacGinitie, W. & Holden, M. Children's conception of word boundaries in speech and print. Journal of Educational Psychology, 1972, 83, 551-557.

Martin, J. & Sakiey, E. Learning sight words by combining common syllable clusters: A case study. Glassboro, NJ: Glassboro State College, 1979. (ERIC Document Reproduction Service No. ED 184 088)

Osburn, W. J. Teaching spelling by teaching syllables and root words. Elementary School Journal, 1954, 55, 32-41.

Otto, W. & Chester, R. Sight words for beginning readers. Journal of Educational Research, 1972, 65, 435-443.

Rinsland, H. D. A basic vocabulary of elementary school children. New York: Macmillan, 1945.

Sakiey, E. The commonest syllables. In M. L. Kamil & A. J. Moe (Eds.), Reading Research: Studies and applications. Clemson, SC: National Reading Conference, 1979.

Sakiey, E. & Fry, E. 3000 instant words. Providence, RI: Jamestown Publishers, 1979.

Sakiey, E. Fry, E., Goss, A. & Loigman, B. A syllable frequency count. Visible Language, in press.

Sakiey, E. & Martin, J. Primary level graphemic syllable lists with pronunciation variations. Paper presented at the meeting of the College Reading Association, Baltimore, October 1980.

Smith, F. & Miller, G. (Eds.) The genesis of language. Cambridge: MIT Press, 1966.

Thorndike, E. L. & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College Press, Columbia University, 1944.

Venezsky, R. L. & Calfee, R. C. The reading competency model. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, DE: International Reading Association, 1970. Pp. 273-291.

Walker, C. M. High frequency word list for grades 3 to 9. Reading Teacher, 1979, 32, 803-812.